BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e25917
doi: 10.3897/biss.2.25917

OPEN ACCESS

Conference Abstract

Using Deep Learning in Collection Management to Reduce the Taxonomist's Workload

Maarten Schermer‡, Laurens Hogeweg§, Max Caspers‡

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ Naturalis Biodiversity Center, Cosmonio Imaging BV, Observation.org, Leiden, Netherlands

Corresponding author: Maarten Schermer (maarten.schermer@naturalis.nl)

Received: 17 Apr 2018 | Published: 15 Jun 2018

Citation: Schermer M, Hogeweg L, Caspers M (2018) Using Deep Learning in Collection Management to Reduce the Taxonomist's Workload. Biodiversity Information Science and Standards 2: e25917. https://doi.org/10.3897/biss.2.25917

Abstract

The completeness and quality of the information in natural history museum collections is essential to support its use, such as in collection management. Currently, the accuracy of the taxonomic information largely depends on expert provided metadata, such as species identification. At present an increase in the use of digitization techniques coincides with a dwindling of the number of taxonomic specialists, creating a growing backlog in specimen identifications.

We are investigating the role of artificial intelligence for automatic species identification in supporting collection management. When identifying collection specimens, common species are predominantly present, taking up a large amount of the expert's time, who has to deal with a relatively easy, repetitive task. Therefore, one of our aims is to use human expertise where it is most needed, for complex tasks, and use properly validated computational methods for repetitive, less difficult identifications. To this end, we demonstrate the use of automatic species identification in digitization workflows, using deep learning based image recognition.

We investigated potential gains in the identification process of a large digitization project of papered Lepidoptera (>500,000 specimens). In this ongoing project, volunteers unpack, register and photograph the unmounted butterflies and repack them sustainably, still

© Schermer M et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

unmounted. Using only the individual images made by volunteers, taxonomic experts identify the specimens. Considering that the speed of digitization currently exceeds that of identification, a growing backlog of yet-to-be-identified specimens has formed, limiting the speed of publication of this biodiversity information. The test case for image recognition concerns specimens of the families Papilionidae and Lycaenidae, mostly collected in Indonesia.

By allowing the volunteers to provide an automatically generated identification with each image, we enable the taxonomic specialists to quickly validate the more easily identifiable specimens. This reduces their workload, allows them to focus on the more demanding specimens and increases the rate of specimen identification. We demonstrate how to combine computer and human decisions to ensure both high data quality standards and reduction of expert time.

Keywords

automated image recognition, deep learning, specimen identification, Artificial Intelligence (AI)

Presenting author

Maarten Schermer

Presented at

SPNHC, theme: Digitisation and Collections Data (oral presentation)